

# Geo-visual Approach for Spatial Scan Statistics: An Analysis of Dengue Fever Outbreaks in Delhi

Shuchi Mala, Raja Sengupta  
Computer Science and Engineering  
Indraprastha Institute of Information Technology Delhi  
New Delhi, India

**Abstract**— There are very few surveillance systems being used to detect disease outbreaks at present. In disease surveillance system, data related to cases and various risk factors are collected and then the collected data is transformed into meaningful information for effective disease control using statistical analysis tools. Disease outbreaks can be detected but for effective disease control, a visualization approach is required. Without appropriate visualization, it is very difficult to interpret the results of analysis. In this work, a method has been developed for geographical representation of the disease surveillance and response system for early detection of disease outbreaks using SaTScan and open source Geographic Information System software. Maps that combine the geographical location of diseases and clusters to enhance the understanding of results of statistical analysis tool are developed using QGIS library which provides many spatial algorithms and native GIS functions. This library is accessed through PyQGIS and PyQt using Python.

**Keywords**—Disease Surveillance; *p*-value; clusters; statistically significant; outbreaks and visualization

## I. INTRODUCTION

Disease surveillance is a continuous process of collecting information, as well as organizing, analyzing and interpreting the information collected so that the disease outbreaks can be determined and effective actions in disease control may be taken. Outbreaks for diseases can be discovered by monitoring space-time trends of disease occurrences which can highlight changing patterns in risk and help to identify new risk factors. However surveillance datasets are mostly large in size.

Therefore, the availability and performance of software capable of analyzing space-time disease surveillance data on a continuous basis is essential for practical surveillance. At various places disease database for various diseases like cancer, dengue fever, malaria and many more are maintained. Number of new cases is typically added to a disease database daily, monthly or yearly, with the duration depending on the type of disease and the limit of database system.

A traditional disease surveillance mechanism involves reporting of diseases confirmed by labs to local or national organizations of health. This does not normally allow for early detection of new outbreaks. In new surveillance systems, it is easier to find out where the outbreaks will occur and what will be the geographical sizes of these outbreaks. In these systems, clustering is done which tells about number of cases of disease within study area.

Visualization of spatial distribution of the disease over a defined area helps user to analyze clusters easily and efficiently; and users can also detect unusual patterns of disease outbreaks. To study geographical variations of disease risk, the locations of cases are mostly proxies for residential addresses such as pin codes. If individual addresses are available, it is easy to plot the locations of cases on the map. The most common type of map for visualization is Choropleth maps for spatial distribution of disease in the well-defined geographical area using health indicators as occurrence, incidence rate and mortality rate. Choropleth maps usually use color or pattern combinations to show different levels of disease risk associated with each geographical area. These geographical areas are small areas usually defined for administrative purposes, such as counties, zones, wards, colonies, villages, towns and cities. For visualization, various geographical information systems can be used like QGIS which is a cross-platform free and open source desktop geographic information systems application.

In this paper, a method is developed to visualize the results obtained by SaTScan. The method provides an easier way to run SaTScan multiple times and add graphical output for analyzing results obtained by the developed application. This standalone package takes datasets or files of population, cases information, geographical coordinates of each location and optional controls as an input in a simple prescribed format and generates text files in SaTScan format and allows the user to choose SaTScan analysis options. It reads the results from SaTScan and creates geographical outputs, based on a separate map boundary file. The front end was developed using PyQt4, Qt4, PyQGIS and QGIS libraries, with Python as the interfacing programming language. Effectiveness of the user interface is demonstrated by a case study of dengue fever in Delhi, India from 2010-2012. This application is particularly useful to health Officials who do not have knowledge of GIS software in disease surveillance and disease control.

## II. MATERIALS AND METHODS

### A. Related work

Disease clustering can be classified as temporal clustering, spatial clustering or space time clustering. Temporal clustering observes whether cases are located close to each other in time, spatial clustering observes whether cases are located close to each other in space and space-time clustering observes whether cases are close in space as well as in time. For detection of disease cluster, various statistical methods have

also been developed. GAM [19] is a cluster detection approach which performs examination of a large number of overlapping circles at a variety of scales and assesses the statistical probability of the number of events occurring by chance. The drawbacks of this method are that it has a multiple testing problems and is heavily computer intensive. FleXScan [23] is free software which was developed to analyze spatial count data using the flexible spatial scan statistic and circular spatial scan statistic. Current version of FleXScan is still restricted specifically to spatial analyses, ignoring the temporal component. Another software Splancs was developed by Rowlingson and Diggle [21] for spatial and space time point pattern analysis. Kulldorff [13] together with Information Management Services Inc. developed SaTScan software that can perform geographical surveillance of a disease, detect clusters and test whether these clusters are statistically significant or not. It can also perform time-periodic disease surveillance for early detection of disease outbreaks. Unlike techniques such as Openshaw's GAM, SaTScan does not take into account the problem of multiple testing and reports the significance of each detected cluster. In the analysis of cluster detection test, there can be issues concerning data, the scale of analysis, correction for covariates and the underlying background population. Covariate in data is the most important problem affecting cluster study. Any spatial or temporal variation of covariates like gender, age, ethnicity, diet, smoking behavior or population density can worsen the real disease patterns. For example, people of similar ethnic origin traditionally tend to live close together, although in present world this is decreasing due to increased population migration. Some diseases can be inherited and if spatial clusters of genetic diseases are need to be observed then examining clusters for such diseases requires evidence for clustering over background population after adjustments made for the genetic covariates. Methods like SaTScan can do such adjustments in covariates. Correction of spatial variations in the population at risk is an important part of spatial epidemiological research because any observed pattern of health events needs to be adjusted by the background population distribution.

There are two issues related to SaTScan software itself. Those are: (1) Lack of cartographic support for interpreting the detected geographical clusters. (2) Outcomes being sensitive to parameter choices related to cluster scaling.

The software does not directly provide any visualization support. Chen et al. [9] suggested that the Geovisual analytics method make it efficient for users to understand the SaTScan results. Authors have illustrated Geovisual analytics approach in a case study analysis of cervical cancer mortality in the U.S. between 2000 and 2004. For all the counties Standardized Mortality Ratio and reliability scores are visualized to identify stable and homogeneous clusters. The proposed Geovisual analytics approach is implemented in Java-based Visual Inquiry Toolkit

### B. Methodology

. In the Field of public health, spatial clustering analysis and subsequent geoprocessing of clustering results is the most efficient yet technically comprehensible way. Kulldorff [13]

described a statistical method for the detection of multi-dimensional point process using spatial scan statistics. It uses variable window size and a baseline process as an inhomogeneous Poisson process or Bernoulli Process. Scanning window can be any predefined shape and is modelled on a geographical space. Monte Carlo sampling is done in which a regular or irregular grid of centroids covering the whole study region is created and then an infinite number of circles around each centroid are created. Actual and expected number of cases inside and outside the circle is obtained and Likelihood Function is calculated. By using Monte Carlo Simulation random replicas of the data set are generated under the null-hypothesis of no cluster. Likelihood function value is ranked with the maximum likelihood ratio from the Monte Carlo replications. These ranks are called p-values. For a cluster to be statistically significant its p-value should be less than one. The cluster with the smallest p-value is the most likely cluster that has occurred not by chance. SaTScan software [8], offers many advantages. It is robust, computationally efficient, has flexibility of options, corrects for multiple comparisons, adjusts for heterogeneous population densities among the different areas in the study, detects and identifies the location of the clusters without prior specification of their suspected location or size thereby overcoming pre-selection bias, and allows for adjustment for covariates. However, one of the drawbacks of SaTScan is that it does not have a visualization system for presenting the results. In this regard GIS has sophisticated mechanisms to visualize data. In addition to being able to assess disease cases with a general categorical definition of "place", GIS systems provide the means to analyse spatial-temporal relationships between sets of variables, allowed users to identify spatial patterns in data, and provided the means to integrate databases on the basis of geography. But it is a time-consuming process and one has to learn how to work with GIS packages. The relationships between software components of the developed standalone application developed in python; termed "Visual Interpretation of Statistical Analysis" (VISA) is shown in Fig. 1. SaTScan is embedded in this application for statistical analysis. The portion in the dotted line can be added to this application for the web enablement. This application is a general one and is applicable to other diseases also.

The standalone package provides maps that combine the geographical location of diseases and clusters to enhance the understanding of SaTScan results. A user does not require knowledge of GIS packages. GIS support is provided by QGIS library which provides many spatial algorithms and native GIS functions.

This library is accessed through PyQGIS by Python bindings which provides simpler programming environment. For developing GUI, PyQt4, Qt4-devel, Qt4-doc and Qt4-libs have been used. PyQt is a Python binding of the cross-platform GUI toolkit Qt. Libraries of QGIS (Quantum GIS) is a collection of C++ classes which can be used for accessing and manipulating spatial objects. The libraries used are Core containing GIS functions, GUI containing controls and User interface such as canvas map which displays and manipulates the maps. ftools, a python plug-in of QGIS is used for spatial analysis i.e. for drawing clusters of specified radius.

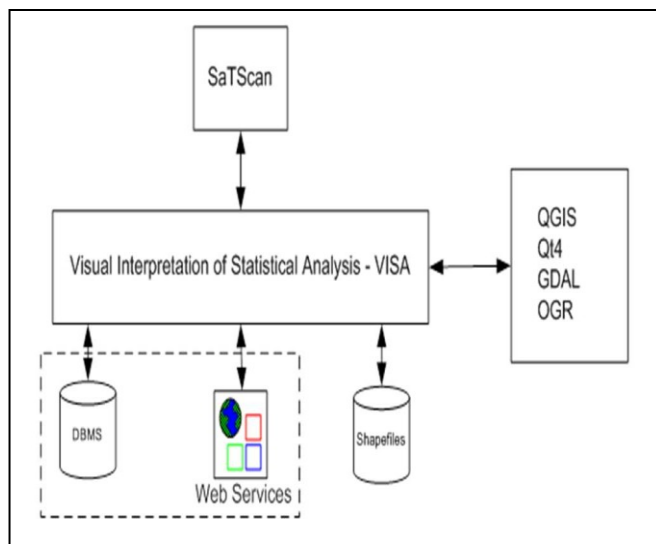


Fig. 1. Structure of Standalone Application VISA

The user interface is developed using Qt Designer and shapefile viewer is developed in python using libraries of QGIS. The first screen generated by the application is shown in Fig. 2. Data regarding number of years of study is read from a text file yeardata.txt which is modified according to the study period. Fig. 2 shows the first screen of the developed application which is generated from yeardata.txt. It shows the buttons for running SaTScan and for generation of disease cases maps and disease density maps for study period.

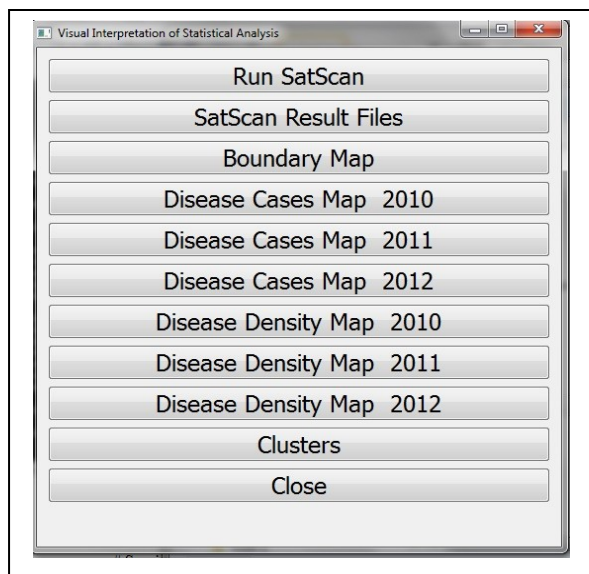


Fig. 2. First screen of the application

For analysis in SaTScan two main files are needed. If either is unavailable then it is not possible to run SaTScan.

These files are: case file and coordinate file. Case file contains information about number of cases, time at which case was reported and covariates like age, gender etc. Coordinate file contains name of location and its longitude and latitude or Cartesian coordinates.

These two files are stored in a folder. Various steps of the method are explained below:

1) When Main Menu appears click "Run SaTScan" button.

2) Now SaTScan menu will be displayed. Select new session. In the Input menu, case file and coordinate files are selected. Study period is entered. In the Analysis menu, type of analysis and probability model is selected. Under Advanced tab, a Circle with radius is selected and value of radius is entered. In Output menu name of result file is entered and additional output files are marked.

3) Results are stored in four output files. For example in the presented work, analysis is done on dengue fever outbreaks. Consequently, the four output files generated are: dengue\_cases.sci.txt, dengue\_cases.llr.txt, dengue\_cases.gis.txt and dengue\_cases.col.txt. The output files are stored in the same folder. When SaTScan Result File is clicked, a name of file with extension col.txt is entered in which SaTScan results are stored.

4) On clicking Boundary Map, shapefile of state or county or country map is imported in the above program as layer and displayed. Since the analysis of dengue cases in Delhi is done in this study, Delhi Map with Zones is displayed as shown in Fig. 3 which contains tool boxes for display Map, ZoomIn, ZoomOut, Pan and zoomfull toolboxes.

5) On clicking Disease Cases Map 2010, first of all, function AppLatLonCase.py is called which extract cases of 2010 and appends longitude and latitude values for each case. A separate file case lt2010.csv is generated. This csv file is read in the program and the layer showing cases which occurred in the year is 2010 is generated and displayed as shown in Fig. 4.

6) On clicking Disease Cases Map 2011 and Disease Cases Map 2012, the operations mentioned in step (v) are repeated for cases occurring in the years 2011 and 2012 as shown in Fig. 5 and Fig. 6 respectively.

7) On selecting Disease Density Map 2010, all the dengue cases in the year 2010 in Delhi are considered. A program computepoint.py is called to calculate number of dengue cases in each zone, a shape file is created and displayed which is shown in Fig. 7. Number of cases which occurred in the year 2010 in each zone is also displayed.

8) On clicking Disease Density Map 2011 and Disease Density Map 2012, the operations mentioned in step vii.

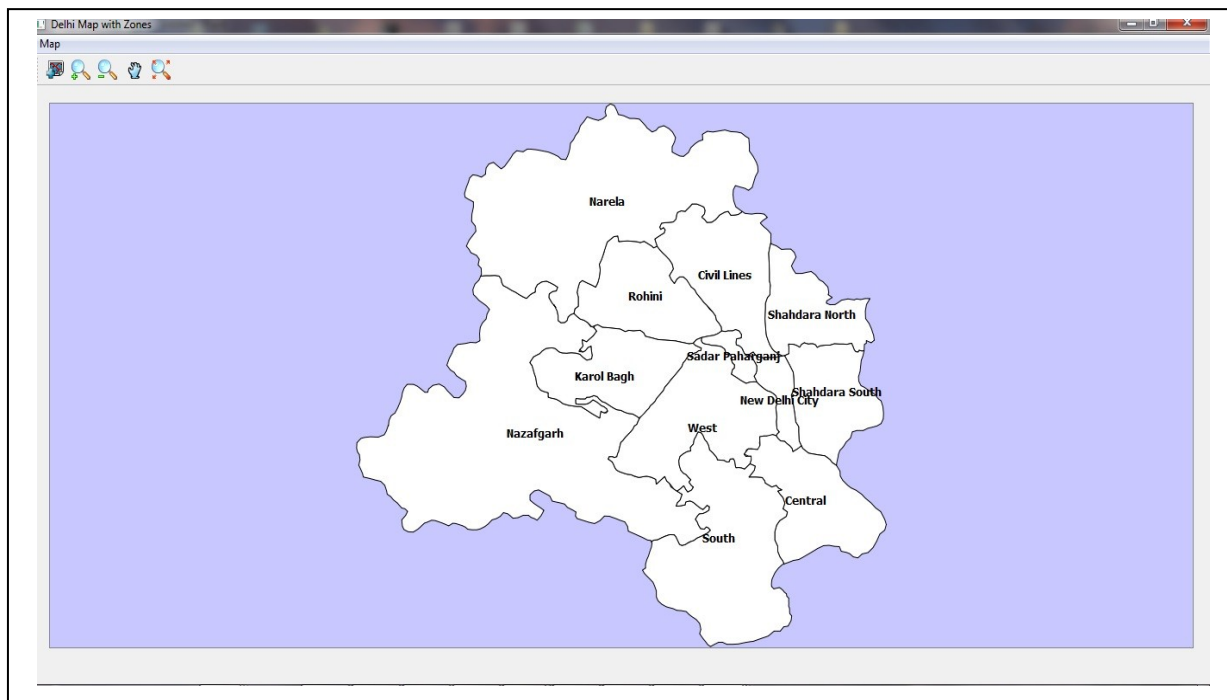


Fig. 3. Screenshot of the District Boundary Map of Delhi

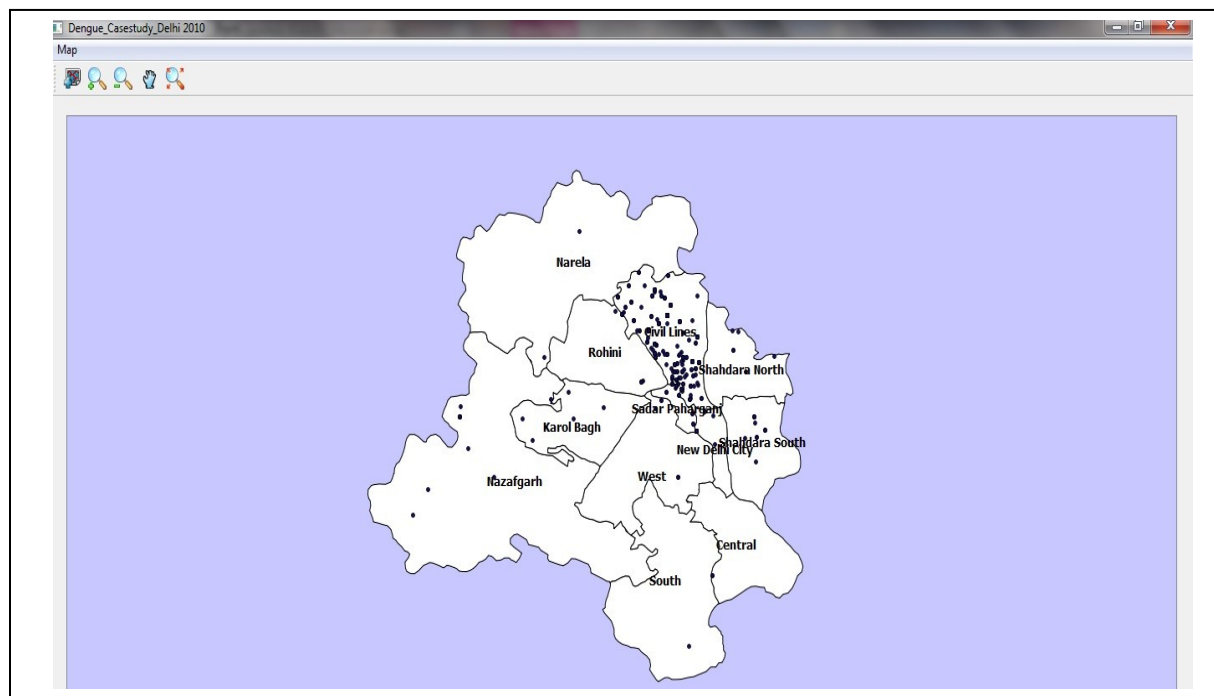


Fig. 4. Dengue cases in Delhi in the year 2010

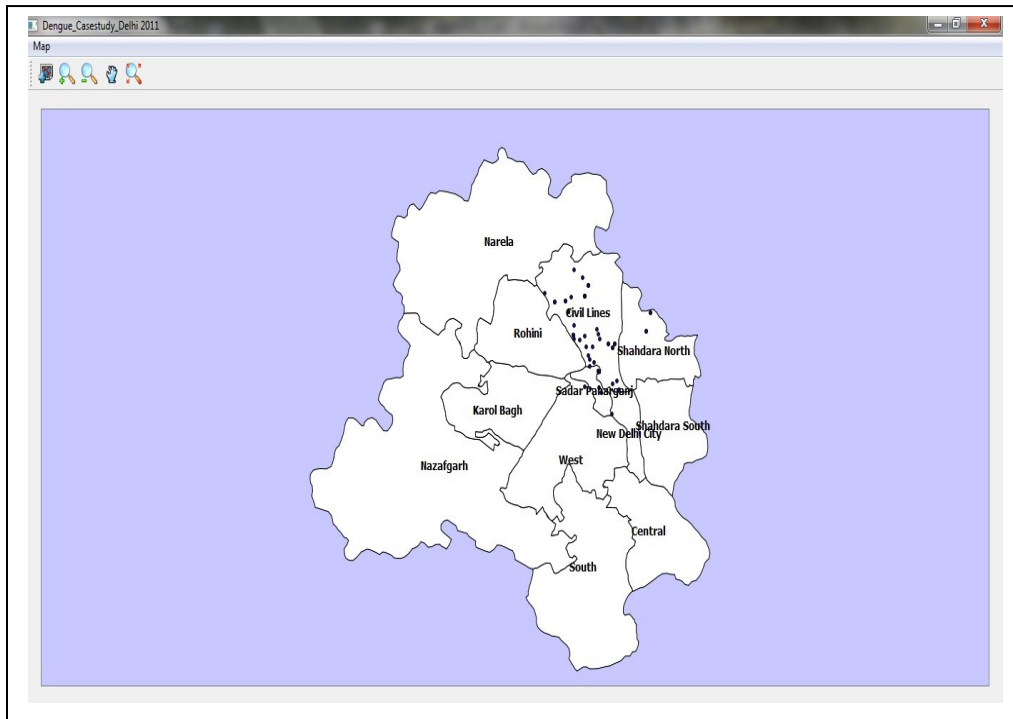


Fig. 5. Dengue cases in Delhi in the year 2011

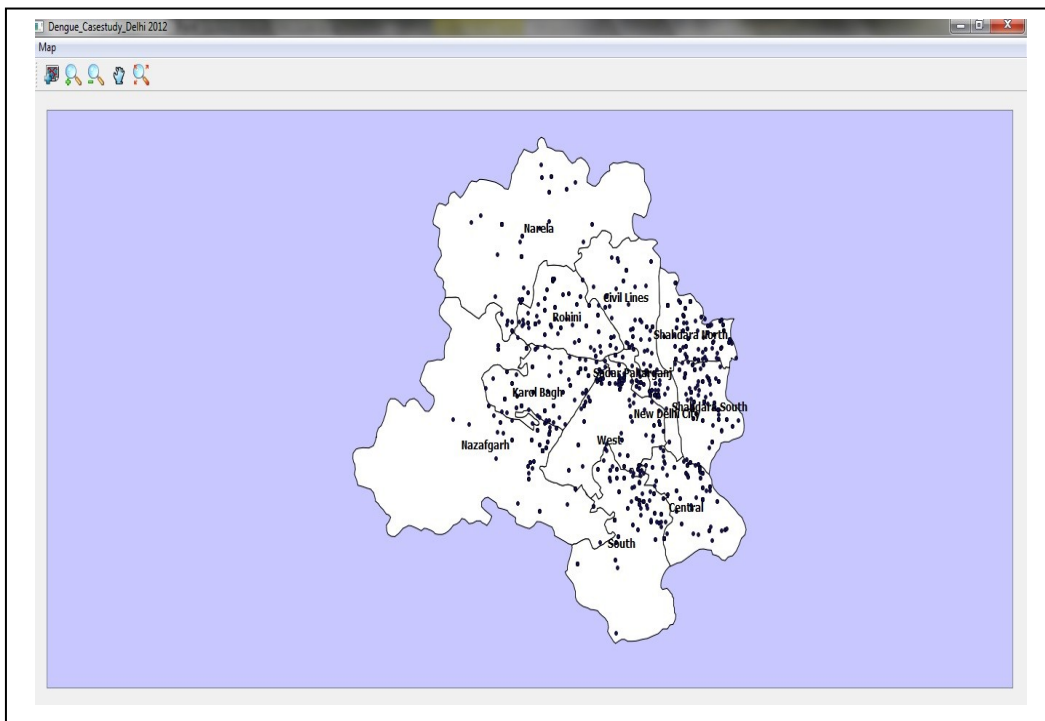


Fig. 6. Dengue cases in Delhi in the year 2012

are repeated for cases occurring in the years 2011 and 2012 as shown in Fig. 8 and Fig. 9 respectively.

1) When Cluster button is clicked, the application reads the SaTScan result file *dengue\_cases.col.txt* and forms a file *cluster.csv* which contains all clusters having *p*-value less than a specified value. This *csv* file is read and a layer is created. Using *fools* plug-in, buffers of a specified radius are created. The created layers are displayed which are shown in Fig. 10.

2) On clicking close, the application is terminated.

### III. RESULTS AND DISCUSSION

In this section an analysis of dengue fever outbreaks in Delhi for past three years is presented as a sample case. In India 28,000 dengue cases were reported in 2010 and there is a significant rise in the incidence of dengue from 18,860 cases in the year 2011 to 49,606 cases in the year 2012 and 1,700 dengue cases were reported from Delhi due to the disease last year. If disease surveillance was done nation-wide for early detection of disease outbreaks, effective actions would have been taken and outbreaks would have been controlled. To perform disease surveillance statistical methods to detect disease clusters are required. It is equally important to have effective visualization approach. The presented work covers both the aspects of disease surveillance along with the surveillance results.

#### A. Data collection

To perform any type of analysis the most importance requirement is data. Data for analysis was collected from DHO Civil Lines Zone, Municipal Corporation of Delhi, Health Department. The data covering details on dengue cases in Delhi are available for the years 2010, 2011 and 2012 only. The details were taken from the Epidemiological Investigation form for the above mentioned years.

#### B. Data Pre-processing

The data collected is transformed in the format that is required to perform statistical analysis. The data is transformed into a comma separated file which contains information about the observed cases location name, case count, date of reported case, age of the case and gender. In this file age and gender are the covariates. Another comma separated value file is created for the coordinate information about the geographical location where case has occurred. In this file information about location name, latitude and longitude of the location of the case is stored.

#### C. Map Digitization

Georeferencing process is used to assign real-world coordinates to each pixel of the raster using QGIS. In the presented work, scanned map of Delhi is digitized by obtaining coordinates from the markings on the map image itself. Using these GCPs (Ground Control Points), the image is warped and it is made to fit within the chosen coordinate system [15].

#### D. Statistical Analysis

To detect clusters of disease outbreaks, statistical analysis is performed using SaTScan software which is embedded in the stand alone VISA application. To begin analysis, in the

GUI of the presented application, click on "Run SaTScan" button. On clicking the button, SaTScan window opens. In the input tab, import the created comma separated value file which contains the details about observed cases under case file option. This file is saved as *dengue\_cases.cas*. The extension supported by SaTScan for case file is *.cas*. Under coordinate file option import the other comma separated file which contains latitude and longitude information about each location of the observed case. This file is saved as *dengue\_cases.geo*. The extension supported by SaTScan for coordinate file is *.geo*. The study period is from 2010/1/1 to 2012/12/31 with time precision to be year and coordinates latitude/longitude. In the analysis, space-time analysis and space time permutation model are selected. Space time permutation model is used in the analysis because dengue fever has a relation with environmental variables as many cases are observed in months with warm and humid weather. Hence with geographical location time is also an important parameter to perform analysis. Adjustment for the maximum spatial cluster size is set at 2 kilometers. In the output tab, result file *dengue cases.txt* is saved and SaTScan software is executed.

#### E. Visualization results generated by VISA package

Delhi region is divided into 12 zones which are Narela, Rohini, Civil Lines, Shahdara North, Shahdara South, New Delhi City, Karol Bagh, Nazafgarh, central, west, Sadar Paharganj and south. On clicking Delhi Map button, the zone boundary map is generated as shown in Fig. 3. On Clicking Disease Cases Map 2010 button, dengue fever cases spread in year 2010 map are generated as shown in Fig. 4. High number of cases occurred in Sant Nagar (64), Jahangirpuri (47), Timarpuri (24) and Burari Village (22). On Clicking Disease Cases Map 2011 button, dengue fever cases reported year 2011 map are generated as shown in Fig. 5. Majority of cases occurred in Jahangirpuri (12), SantNagar (5), and Malkaganj (4). On Clicking Disease Cases Map2012 button, dengue fever spread in year 2012 map is generated as shown in Fig. 6. High number of cases occurred in Rajnagar-II (34), Mandawali (26) and SangamVihar (23). There are 8 locations in Delhi where cases were reported in 2010, 2011 as well as 2012 as TABLE 1. On clicking "No. of cases Map" button, Choropleth map is generated to show density of dengue cases with graduated colors. On the basis of zone, most cases are reported from Civil Lines Zone, Shahdara North Zone and Shahdara South Zone as shown in Fig. 6. On clicking Clusters button, clusters and buffer map are generated as shown in Fig. 7. There are seven detected clusters out of which one is most likely cluster and six are secondary clusters as shown in TABLE 2.

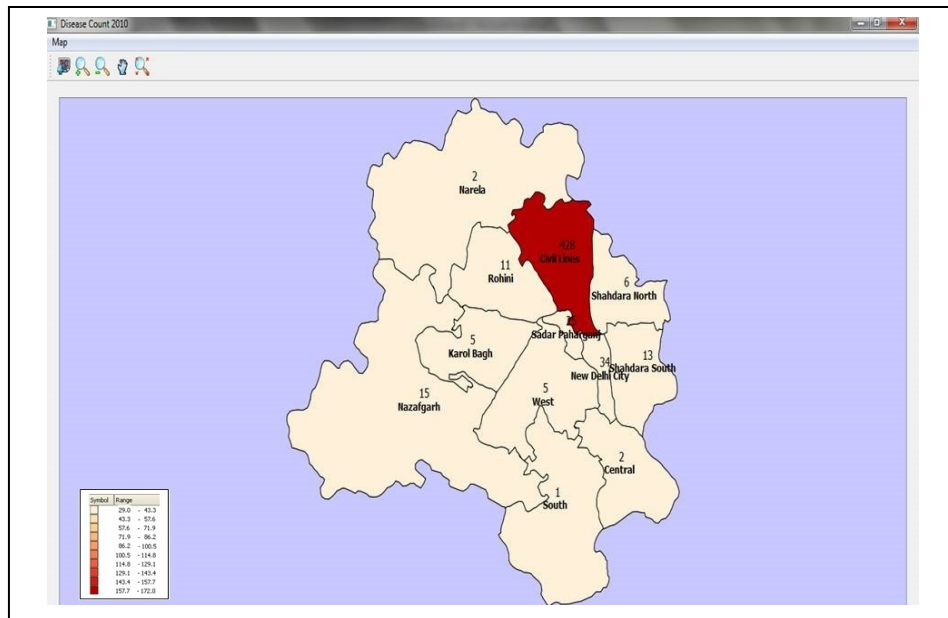


Fig. 7. Dengue cases in 2010 is shown by graduated colour

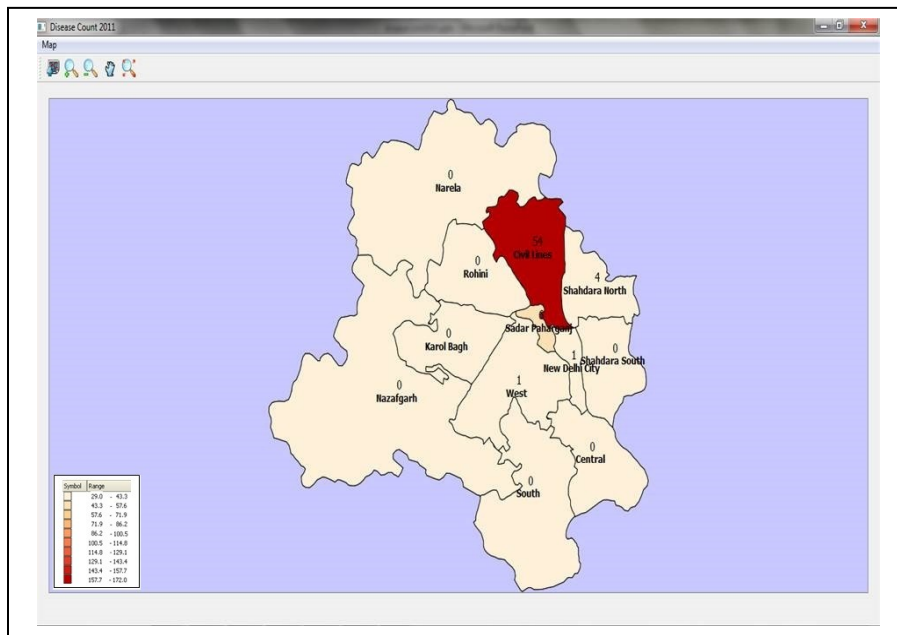


Fig. 8. Dengue cases in 2011 is shown by graduated colour

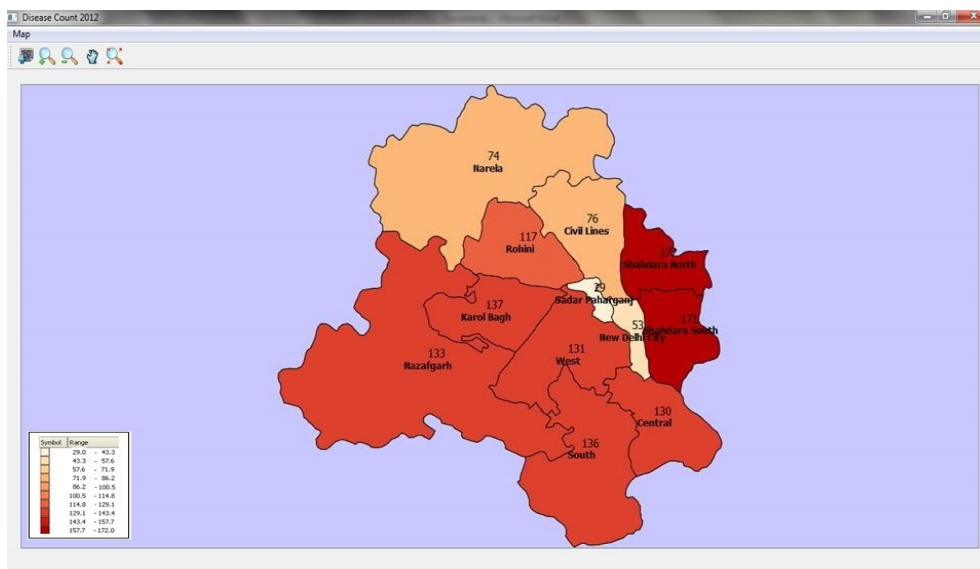


Fig. 9. Dengue cases in 2012 is shown by graduated colour

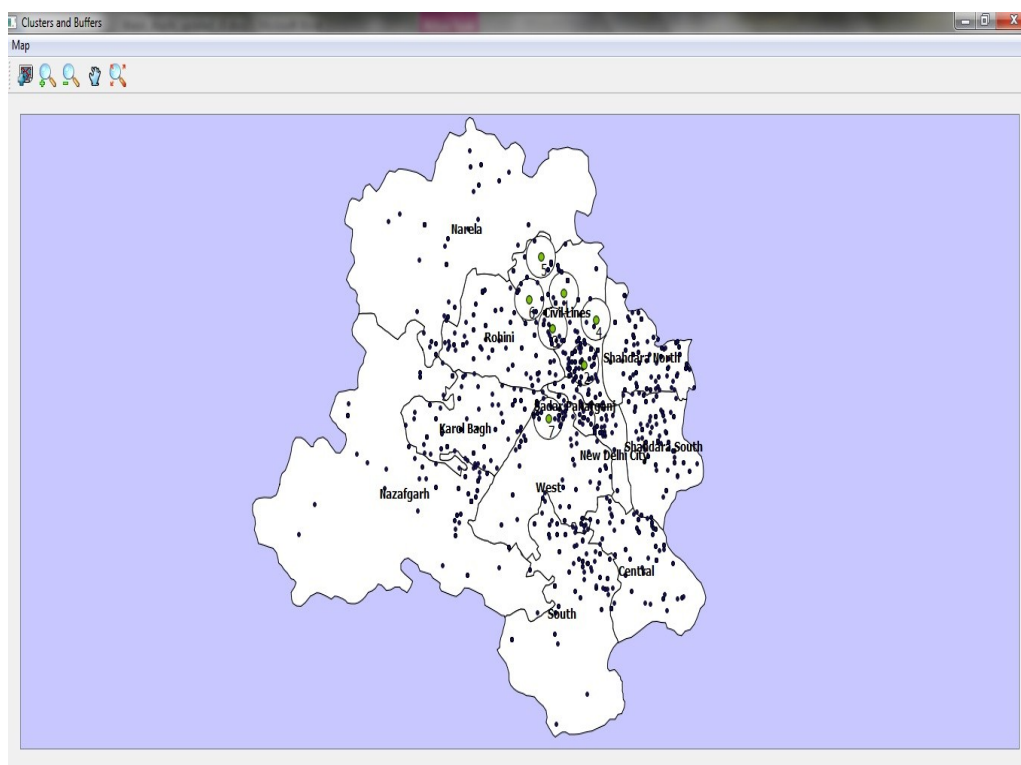


Fig. 10. Buffer zones of 2 km radius around the clusters



F. Interpretation of detected clusters

Total number of locations where cases were reported is 617 and total number of cases reported is 1976. The shape file of the clusters generated and saved by the package is easily imported as a KML file in Google Earth so that interpretations can be made easily. The following are the details of detected clusters:

1) The locations included in the most likely cluster are Sant Nagar, Burari; Sant Bhagat Colony, Burari; Burari Village; Mukandpur; Radha Vihar; Bengali Colony; Bhalswa and Kaushik Enclave with number of cases observed to be 119 and expected number of cases as 48.16 with test statistic as  $38.097$  approximately, highest among detected clusters. It is statistically significant because its  $P$ -value is smaller than  $0.10 \times 10^{-16}$  which shows that its occurrence is not by chance. The reasons for having high number of outbreaks in the regions related to this cluster are as follows:

- Burari region has Jharoda Dairy where cattle are raised. Mostly, dairies use earthen pots which are not cleaned frequently and results in stagnant water, helping in breeding of mosquitoes.
- Bhalswa region has Bhalswa Lake and there is a high probability that the water present in the lake is stagnant.
- Massive construction sites with some unregulated construction sites.
- Demographic in this cluster is lower middle class and migrant population.
- During rainfall Yamuna bank is flooded and there is no place to drain water from these places.

TABLE 1. LOCATIONS WITH NUMBER OF CASES (2010-2012)

Location	Total cases count		
	2010(cases count)	2011(cases count)	2012(cases count)
Timarpur	24	2	5
Sant Nagar	62	5	3
Nehru Vihar	2	3	3
Manju Ka Tilla	19	6	2
Kingsway Camp	10	1	1
Kamla Nagar	7	1	1
Jahangirpuri	47	12	12
Azadpuri	16	3	1

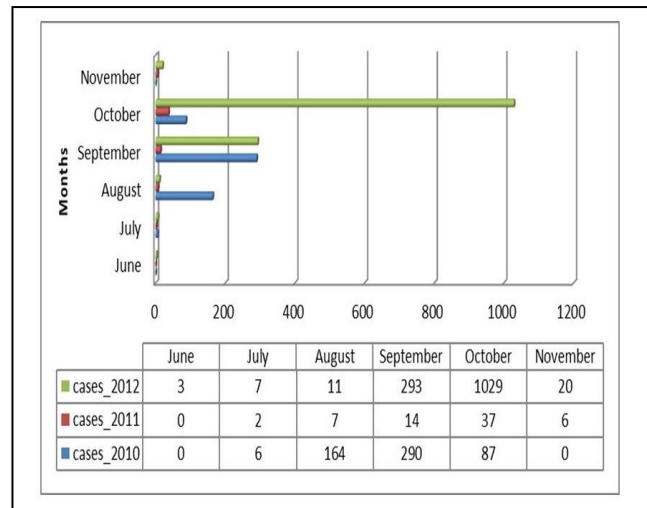


Fig. 11. Number of cases on the basis of months

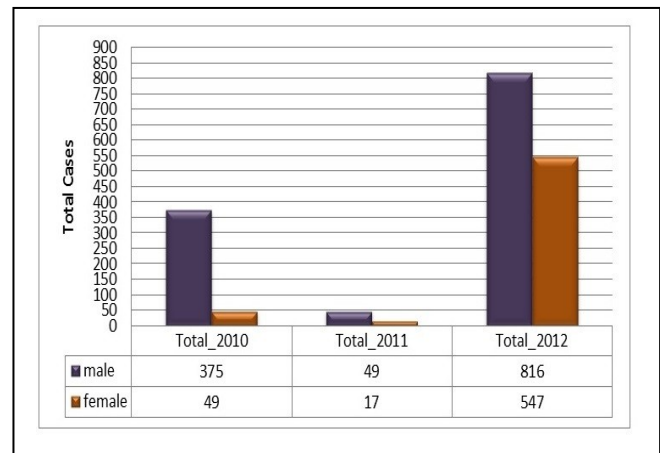


Fig. 12. Number of cases on the basis of gender

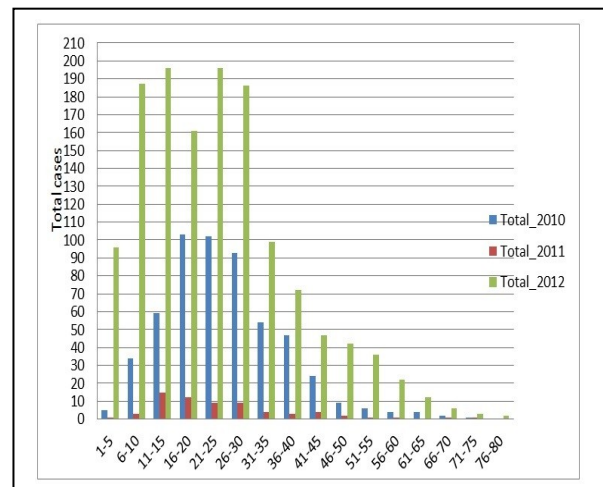


Fig. 13. Number of cases on the basis of age groups

TABLE 2. CLUSTERS WITH P-VALUE LESS THAN 0.55

Cluster Number	Clusters Information									
	Central Location ID	Latitude	Longitude	Cluster Start Date	Cluster End Date	Log Likelihood Ratio	P-value of Cluster	Observed Cases	Expected Cases	Observed/Expected
1	Sant Nagar Burari	28.74544	77.190212	01-01-2010	31-12-2010	37.490136	1.00E-17	119	48.85	2.45
2	Delhi University	28.7	77.213	01-01-2010	31-12-2010	30.182756	1.11E-16	142	69.66	2.04
3	Majlisparki	28.7	77.177633	01-01-2010	31-12-2010	24.49099	2.7E-13	100	45.94	2.18
4	Wazirabad Village	28.7	77.226145	01-01-2010	31-12-2010	8.08196	0.0022	22	8.13	2.71
5	Kadi Village	28.8	77.164402	01-01-2010	31-12-2010	8.03457	0.0023	24	9.35	2.57
6	Badli	28.7	77.151392	01-01-2011	31-12-2010	5.921857	0.042	6	0.97	6.2
7	East Patel Nagar	28.6	77.172961	01-01-2012	31-12-2010	3.848719	0.548	65	45.35	1.43

2) The locations included in the second cluster are Delhi University, St. Stephen College, Patel Chest, Miranda House, Ramjas College, Mall Road, Old Vidhan Sabha, Mourice Nagar, Vijay Nagar, Khyber Pass, Hansraj College, Hakikat Nagar, Old Chandrawal, Kamla Nagar, Rajpur Road, Old Gupta Colony, New Chandrawal, Under Hill Road, IP College Campus, Timarpur, Outram Lane, Aruna Nagar, Civil Line, Out Ram Line, Kalyan Vihar, Malkaganj, Rajpur, Kingsway Camp, Ghanta Ghar, Gur Mandi, GTB Hospital Campus, Ram Kishore Marg, CC Colony, Kabir Basti, Ishwar Colony, Singh Sabha Road, Indra Vihar, Rana Pratap Bagh, Mukherji Nagar, West Mukherji Nagar, Shakti Nagar and Majnu Ka Tilla with number of cases observed as 142 and expected number of cases as 68.82 with test statistic to be 31.074 approximately, second highest among detected clusters. It is statistically significant because its P-value is  $0.11 \times 10^{-15}$  which shows that its occurrence is not by chance. The reasons for having high number of outbreaks in the regions related to this cluster are as follows:

- Most of the areas are institutional and government office area. These are vulnerable due to lack of awareness.
- There is no sense of ownership because of multiplicity of authorities.
- There are large open areas near Rajpur Road where rain water is collected
- Large student population and lack of awareness about prevention control of disease.

In months of July and August breeding of mosquitoes takes place and in months of September and October they become adults and thus large number of cases are observed in these two months as shown in Fig. 11.

On the basis of gender, infected male population was very high in 2010 and 2011 in comparison to female population but

in 2012 there was a high increase in the number of infected females as shown in Fig. 12.

On the basis of age group in the year 2010, most cases occurred in 16-20 years of age; in the year 2011 most of the cases occurred in the 11-15 years of age and in the year 2012 most of cases occurred in the 11-15 years and 20-25 years of age groups as shown in Fig. 13.

#### IV. CONCLUSION

The Geographical visualization approach developed in this paper facilitates the space-time cluster detection methods by providing an efficient representation of the results of statistical analysis in geographical space. Space-time analysis is performed by using space-time permutation model of SaTScan. The disease clusters are detected with the cluster radius as 2 kilometers. With the proposed visualization method maps are generated. These maps show spreading of cases, density of cases within each district or county and statistically significant clusters. With the help of presented work proactive actions can be taken to prevent disease outbreaks. It is also helpful in identifying the hot zone of an epidemic. Therefore on the basis of information gathered from statistical analysis and visualization the overall quality of health of the nation can be improved. This application is specifically useful to health Officials who do not have knowledge of GIS software. The standalone package is developed by using python, PyQGIS, QGIS, PyQt4, Qt Designer and SaTScan. Use of the proposed method is demonstrated to analyze the results of statistical analysis of dengue fever. The team of doctors in Municipal Corporation of Delhi found these results very informative; efficient and accurate and the interpretations related to the clusters were very similar to their interpretations.

#### FUTURE WORK

The present work will be extended to include a statistical interface for space-time analysis instead of integrating the developed application with SaTScan. This application will be

made web-based and database containing diseases information will be integrated with it. With the facility of a database user can save data whenever required and can perform analysis at chosen time intervals such as on weekly basis or based on months or years from remote location.

#### ACKNOWLEDGMENT

We would like to thank Dr Ashok Rawat, Deputy Health Officer, Civil Lines Zone, Municipal Corporation of Delhi and Dr. Sanjay Sinha, Senior Doctor, Health Department, Civil Lines Zone, Municipal Corporation of Delhi for providing us the data on dengue fever to perform analysis.

#### REFERENCES

- [1] Agricola Odoi, S Wayne Martin, Pascal Michel, Dean Middleton, John Holt and Jeff Wilson, 2004, Investigation Of Clusters Of Giardiasis using GIS And A Spatial Scan Statistic, International Journal of Health Geographics, Vol.3, No.11, pp 1-11.
- [2] Allyson M Abrams and Ken P Kleinman, 2007, A SATSCAN Macro Accessory For Cartography (Smac) Package Implemented With SAS Software, International Journal of Health Geographics, Vol. 6, No.6, pp. 1-8.
- [3] Besag J.E. and Newell J., 1991, the detection of clusters in rare diseases, Journal of the Royal Statistical Society, Series a, 154, 143–155.
- [4] Sabel E. Clive, Loytonen Markku, Maheswaran Ravi, Craglia Massimo, 2004, GIS in public health practice, CRC Press LLC.
- [5] Tatlonghari Jayson A. Rick, 2009, GIS Manual using Quantum GIS, pp. 23-36.
- [6] Dobias Martin, 2010, PyQGIS Developer Cookbook, created by sphinx-quickstart.
- [7] B. M. Harwani, 2011, Introduction to Python Programming and Developing GUI Applications with PyQt, Cengage Learning, pp.2 01-416.
- [8] Kulldorff M., 2011, SaTScan User Guide, pp. 4-107.
- [9] Chen Jin, Roth E Robert, Naito T Adam, Lengerich J Eugene and MacEachren M Alan, 2008, Geovisual analytics To Enhance Spatial Scan Statistic Interpretation: An Analysis Of U.S. Cervical Cancer Mortality, International Journal of Health Geographics, Vol.7, No.57, pp. 1-18.
- [10] Kulldorff M., 1998, Statistical methods for spatial epidemiology: tests for randomness, In GIS and Health, pp. 49–62, London: Taylor & Francis.
- [11] Gail M., Krickeberg K., Samet J., Tsiatis A., Wong W., Tango. T, 2009, Statistics for Biology and Health, Springer.
- [12] Jankowska Marta, Aldstadt Jared, Getis Arthur, Weeks John, Fraley Grant, 2008, An Amoeba procedure For Visualizing Clusters, GIScience, Park City, UT.
- [13] Kulldorff M., 1997, A Spatial Scan Statistic, Commun Statist Theory Meth, Vol.26, No.6, pp. 1481-1495.
- [14] Kulldorff M., 2001, Prospective Time Periodic Geographical Disease Surveillance Using A Scan Statistic, Royal Statistical Society, Vol. 164, Part 1, pp. 61-72.
- [15] Kulldorff M., Feuer J. Eric, Miller A. Barry and Freedman S. Laurence, 1997, Breast Cancer Clusters in the Northeast United States: A Geographic Analysis, American Journal of Epidemiology, Vol. 146, No. 2, pp. 161-170.
- [16] Kulldorff M., Huang L., Pickle L. and Duczmal L., 2006, An Elliptic Spatial Scan Statistics, Wiley InterScience, Vol. 25, pp. 3929-3943.
- [17] Kulldorff M., Heffernan Richard, Hartman Jessica, Assunc Renato, Mostashari Farzad, 2005, A Space Time Permutation Scan Statistic For Disease Outbreak Detection, plos journals, Vol.2, No.3, pp. 216-224.
- [18] Kulldorff M., Athas F. William, Feuer J. Eric, Miller A. Barry and Key R. Charles, 1998, Evaluating Cluster Alarms: A Space-Time Scan Statistic and Brain Cancer in Los Alamos, New Mexico, American Journal of Public Health, 1998, Vol. 88, No.9, pp. 1377-1380.
- [19] Openshaw, Charlton M, Wymer C, Craft A, 1987, A Mark 1 geographical analysis machine for the automated analysis of point data sets, International Journal of Geographical Information Systems, Vol.1, pp. 335–358.
- [20] Piarroux Renaud, Barrais Robert, Faucher Benot, Haus Rachel, Piarroux Martine, Gaudart Jean, Magloire Roc, and Raoult Didier, 2011, Understanding The Cholera Epidemic, Haiti. Emerging Infectious Diseases, Vol. 17, No. 7, pp. 1161-1167.
- [21] Rowlingson B. and Diggle P., 1993, Splanx: spatial point pattern analysis code in SPlus, Computers and Geosciences, Vol. 19, pp. 627–655.
- [22] Sheehan TJ and DeChello LM, 2005, A Space-Time Analysis of the Proportion of Late Stage Breast Cancer in Massachusetts, 1988 to 1997, International Journal of Health Geographics, Vol.4, No.15, pp. 1-9.
- [23] Takahashi K., Yokoyama T., and Tango T., 2009, FleXScan: Software for the Flexible Spatial Scan Statistic. v3.0, National Institute of Public Health, Japan.
- [24] Tango T and Takahashi K, 2005, A Flexibly Shaped Spatial Scan Statistic for Detecting Clusters, International Journal of Health Geographics, Vol. 4, No. 11, pp. 1-15.